REPORT DOCUMENTATION PAGE

AFRL-SR-AR-TR-09-0065

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
30-08-2008	Final Report	01-08-05 to 31-05-08

4. TITLE AND SUBTITLE
Research on a mathematical framework and practical applications of system architecture

5a. CONTRACT NUMBER
FA9550-05-1-0454

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Peter P. Chen

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Louisiana State University
Tower Road
Baton Rouge, LA 70803

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office of
Scientific Research
875 N. Randolph Dr.
Arlington, VA 22203
DR LUGINBUHL/NL

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
In this project, we have studied several critical issues of system architecture including: a new mathematical framework and operators of system architecture, practical applications of mathematical framework to new system architecture for the next generation of database/information systems and to profiling problems in anti-terrorism and cyber security.

20090324163

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*

ABSTRACT

A system consists of components, which could be systems themselves. Sometimes, different ways of linking the components together may produce systems with different behavior and properties. In the past, many projects were delayed or designed incorrectly because of lack of understanding of system architecture. In building large-scale systems (information systems and non-information systems) in the past twenty years, most organizations (no matter they are in the military/government domain or in the private domains) became recognized the importance of system architecture. In this project, we have studied several critical issues of system architecture including: a new mathematical framework and operators of system architecture, practical applications of mathematical framework to new system architecture for the next generation of database/information systems and to profiling problems in anti-terrorism and cyber security.

.

I. INTRODUCTION

Many large-scale system projects were cancelled or significantly delayed, primarily due to incorrect or the lack of system architectures. Maintenance of existing systems has been a nightmare for similar reasons. An Air Force-LSU System Architecture Workshop was held in FY04 in New Orleans to identify promising areas of research which could help define both current and next generation DoD System Architectures to avoid this kind of problems. Users described various problems with the development of system architectures. These ranged from (1) faulty system descriptions, (2) systems costing too much to operate, (3) inflexibility and time to accommodate mission changes, (4) little interoperability across and with other systems and (5) frustrations with changing architecture requirements and costly system re-developments.

The problem of system architecture outlined at the workshop by Dr Alex Levis, former Chief Scientist for the AF, can be expressed as follows: "The focus of architectures is on information systems that consist of both hardware and software and consequently both approaches are needed. However, the approaches are not fully compatible (e.g., dictionaries) causing serious problems in leveraging architecture efforts. In addition it takes a long time to adopt an architecture-based approach in acquisition and in design. The problem in the short term is inadequate experience with the architecture approach, insufficient training, and a serious deficiency in capability to evaluate architectures and to compare alternative architectures for the same operational concept. However, there is sufficient theory in mathematics and computer science and modeling and simulation technology that needs to be exploited for this class of problems but a need exists to build on those and develop applications to the desired end.

What is needed are promising new research directed at producing an efficient, architecture-based systems engineering process that enabling the integration of legacy systems with new systems and to exploit technology advances to provide desired capabilities to the war-fighter. As Dr. Joel Moses (MIT) has stated "The importance is to understand the trade-offs between complexity, flexibility, and performance for various architectures." New system development techniques must take advantage of the changing world of science and information technology with views that design of innovative system architectures can be both a scientific and art development with a complete set of supporting tools.

II. A PRELIMINARY MATHEMATICAL FRAMEWORK OF SYSTEM ARCHITECTURE

There are many research issues in the system architecture area. In this project, we have developed a preliminary framework of system architecture:

Mathematical Foundation of System Architecture

It is well recognized that there is a very strong need for more precise definitions and specifications of systems using mathematical (particularly, the algebraic) formalism. In this project, we have proposed and studied a preliminary mathematical framework for system architecture, which is an algebraic system that consists of the following major ingredients:

- Three Types of Sets
  - Entity Sets, Relationship Sets, and Value Sets
  - Relationship is a "mathematical relation" defined on a Cartesian Products of Entity Sets
  - Attributes are "mappings/math. relations"
- A Set of Algebraic Operators
  - Operations on Entity Sets
  - Operations on Relationship Sets
  - Operations on Value Sets
- Innovative/New Features

- Operators have "cost functions," "time duration function," "pre-conditions," "after-conditions," etc.

Examples of the algebraic operations are:

- Composition of relationships:
  - Parents (Parents (x: person))
    = grandparent (x: person)
- Construction of a high-level entity (i.e., assembly) from several low-level entities (components):
  - W = Construct ([x, y, z,…], where x, y, z are … and cost functions, conditions, constraints)
- Deletion of Relationship
- Addition of Relationship
- Move an entity = break up relationship(s) and addition of (new) relationship(s)

We believe this type of analysis will be useful in the analysis and evaluation of alternative system architectures.

III. INVESTIGATION OF PRACTICAL APPLICATIONS OF MATHEMATICAL FRAMEWORK FOR SYSTEM ARCHITECTURE

We have investigated two applications of the mathematical framework: (1) application to organization design, (2) application to a next generation of database/information systems.

III.A. Application to Organization Design and System Architecture

A simplified version of the algebraic operations was investigated and can be used in the analysis of a certain family of tree/hierarchical structures for the study of complexity and the optimal structure of an organization. We introduce a complexity function over all rooted trees (that is, Tree Hierarchies) and studied the complexity of a hierarchy [1]. The purpose of this research is to minimize this function, over all rooted trees on a fixed number of vertices. The results will be useful in the study of organization design and system architectures.

III.A.1. Introduction to Complexity of a Hierarchy and a Rooted Tree

Let us introduce and study a new concept, the *complexity* of a hierarchy.

Many real world structures can be modeled by hierarchies. These examples range from business and/or government organizations to various databases. What we are interested in is how to measure the efficiency of such structures.

Here, we will limit ourselves to tree hierarchies. That is, we only consider hierarchies that can be represented by rooted tress. This restriction does not lose too much generality since tree hierarchies do represent a big percentage of hierarchies used in real world. On the other hand, this restriction does allow us to present a closed form solution.

Before we formally define the complexity of a rooted tree (a tree hierarchy), we point out that there are two factors that increase the complexity. If the tree has a vertex very far away from the root, then the complexity should be high. This is understandable since passing information from the root to this vertex will take a long time, which means the hierarchy is not very efficient. Similarly, if a vertex has too many immediate descendants, communicating with these descendants would easily overwhelm this vertex, which also reduces the efficiency of the hierarchy.

In the next section, we introduce a definition of the complexity of a rooted tree that takes both factors into consideration. Our main result, Theorem 1, says that "near-complete k-ary trees" (will be defined in the next section) are the most efficient rooted trees. Moreover, our proof implies that every rooted tree can be transformed into an efficient tree by repeatedly applying some local minor modifications.

III.A.2. Problem Formulation

Let T be a rooted tree with root r. A vertex v of T is at level ℓ if the unique path from r to v has ℓ edges. The *height* of T, denoted by $h(T)$, is the largest ℓ so that T has a vertex at level ℓ. If $X \subseteq V(T)$ contains r and the restriction of T to X is connected, then we call this restriction a *root subtree* of T, where r is also the root of this subtree.

For each vertex v of T, we denote by $c(v)$ the number of children of v. Let $k \geq 2$ be a fixed positive integer. We call T a k-ary tree if $c(v) \leq k$, for all vertices v of T. A k-ary tree of height h is complete if $c(v)=k$, for every vertex at a level less than h, and $c(v)=0$, for every vertex v at level h. We denote this tree by $T_{k,h}$. A k-ary tree of height h is near-complete if it is obtained from $T_{k,h}$ by deleting some, could be zero, but not all, vertices at level h.

Let $p_k(x)=x-k$, when $x \geq k$ and let $p_k(x)=0$ when $x \leq k$.

Let $f(T)=h(T)+\sum_{v \in V(T)} p_k(c(v))$,
which we call the *complexity function* of T.

Observation 1.
If T is a near-complete k-ary tree on n vertices, then it is straightforward to verify that

$$h(T)=\lceil \log_k (1+n(k-1)) \rceil -1.$$

Theorem 1.
If T is a rooted tree on n vertices, then $f(T) \geq \lceil \log_k (1+n(k-1)) \rceil -1$.

To prove this theorem, we first prove a lemma. Let T be a rooted tree. For i=0,1, let $L_i(T)$ be the set of vertices v with $c(v)=i$.

Lemma 1.
$2|L_0(T)|+|L_1(T)| > |V(T)|$.

Proof.
Let $L_2(T)=V(T)-L_0(T)-L_1(T)$. Since every non-root vertex is a child of a unique vertex,

$$|V(T)|-1=\sum_{v \in V(T)} c(v).$$

Consequently,

$$|L_0(T)|+|L_1(T)|+|L_2(T)| > 2|L_2(T)|+|L_1(T)|,$$

which can be simplified as $|L_0(T)| > |L_2(T)|$, and thus the Lemma follows. ■

Let T be a rooted tree. For each vertex v of T, it is clear that deleting v from T results in exactly $c(v)$ components that contain descendants of v. These components are called *branches* at v. We also define

$$g(T)=\sum\{c(v): v \in V(T) \text{ and } c(v)>k\}.$$

Proof of Theorem 1.
We choose a rooted tree T with the following properties:

(1) f(T) is minimized;
(2) subject to (1), $g(T)$ is minimized;
(3) subject to both (1) and (2), the largest rooted subtree T' of T, such that T' is a near-complete k-ary tree, is maximized.

Clearly, by Observation 1, we need only prove that T is a near-complete k-ary tree, which is equivalent to $T=T'$.

We first prove that T is a k-ary tree. Suppose, on the contrary, that T has a vertex v for which $c(v)=c>k$. Let $T_1, T_2, ..., T_c$ be the branches at v. Without loss of generality, we may assume that $|V(T_1)| \leq |V(T_i)|$, for all i. Let S be the rooted tree, with v as its root, that consists of $T_2, T_3, ..., T_c$ and v. By Lemma 1, and the choice of T_1, we have

$2|L_0(S)|+|L_1(S)| > |V(S)| > |V(T_1)|$.

Now we modify T as follows. First, we delete all edges that are incident with some vertex in $V(T_1)$. Then we add a new edge from each vertex in $V(T_1)$ to a vertex in $L_0(S) \cup L_1(S)$ such that each vertex in $L_0(S)$ is incident with at most two of these new edges and each vertex in $L_1(S)$ is incident with at most one of these new edges. From the above inequality we know that this is possible.

Let T^* be this new tree. Then $h(T^*) \leq h(T)+1$ and v has exactly one child less in T* than it has in T. Moreover, for every other vertex u, either $c(u)$ does not change or u has at most two, with is less than or equal to k, children in T^*. Therefore, $f(T) \geq f(T^*)$ and $g(T)>g(T^*)$, contradicting either (1) or (2). This contradiction proves that T is a k-ary tree.

It remains to prove that $T'=T$. Since T' is a near-complete k-ary tree and it is a rooted subtree of the k-ary tree T, it follows that every vertex of T at level $\ell \leq h(T')$ belongs to T'.

Suppose, on the contrary, that $T' \neq T$. Then T must have a vertex u at level $h(T')+1$. By the maximality of T', adding u to T' does not result in a near-complete k-ary tree. This means that T' is not a complete k-ary tree, which implies that some vertex v of T at level h (T')-1 has fewer than k children. Now we modify T as follows. Take a vertex w at level $\ell > h(T')$ with $c(w)=0$. Delete the only edge incident with w and add a new edge vw. Let R be this new rooted tree. It is clear that $f(T) \geq f(R)$ and $g(T)=g(R)$. However, if we take R' to be the rooted subtree of R that consists of $T'$and w, then R' is a near-complete k-ary tree, which is bigger than T', contradicting (3). This contradiction proves that $T=T'$, which completes the proof of Theorem 1.

III.A.3. Possible Applications and Extensions

The definition of "complexity" in this paper is somewhat different than definitions of complexity by other researchers [2-6]. We think our definition will be useful in the study of optimal organization structure and also the costs of migrating from one organization structure to another.

Another possible extension and application is to develop mathematical framework including a set of algebraic operations, which will be useful in studying different alternatives of systems architectures [7-12].

III.B. Application to the Architecture of Next Generation of Database/Information Systems

We have applied the mathematical framework to a new research area called, "Active Conceptual Modeling of Learning," which incorporate the mathematical framework into the Entity-Relationship (ER) model [10]. After extensive study, we have proposed how to architect a future generation of database/information systems based on the "Active Conceptual Modeling of Learning" [13].

III.B.1. The Challenge

The current information processing technologies can support only "simple rewind." For example, it is possible to use computer backup files to a certain date in the past, and it is also possible to retrieve a previous day's newspaper contents, etc. However, the detailed changes leading to most recent contents were often omitted. Therefore, the existing database and information system technology do not support the learning from the past very well.

The challenge is the development of **next generation of information system** that it will keep and provide detailed changes on records and their corresponding real world environment under which the changes were made. We are proposing to take the features of Active Conceptual Modeling that could capture and represent both the static and dynamic aspects of the real world domain. The persistently stored changes would be organized and interconnected for permitting effective learning from the collection. The traces of past data could be done with multiple dimensions and perspectives so that questions on "who, what, where, when and how" changes occurred could be answered. Significant sequences of events could be extracted from the comprehensive collection for intense review, analysis, and compare so that potentially hidden factors and implicit relationships could be uncovered.

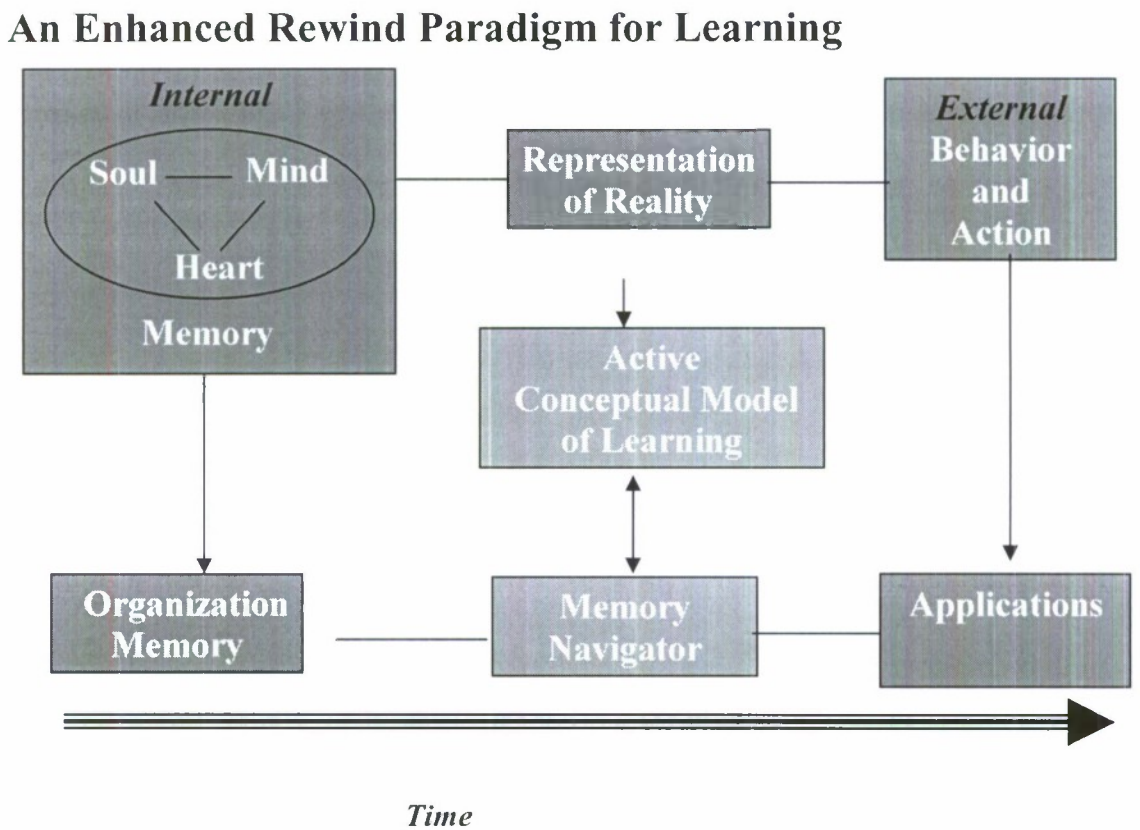
III.B. 2. An Enhanced Rewind Paradigm for Learning

An Enhanced Rewind Paradigm for Learning as one of the approaches in response to the challenge is identified and illustrated. The required features of the paradigm are as follows:

1. A continual process of modeling events and their changes within a given domain.
2. Backtracking of the events to the point and time of interest.
3. Identification of related and parallel events during the period when an incident occurred for possible hidden facts and relationships.
4. Modeling and learning of historical information from different perspectives.
5. Generating and triggering alerts on potentially incidents based on past experiences.

Active Conceptual Model of Learning is incorporated into the Paradigm to ensure the presentation of real world dynamic environment. Multi-level and multi-perspective learning strategies are integrated and included for examining the dynamic collection with different learning logic and approaches. The goals and objectives of planners and decision makers and environmental constraints can be entered into the paradigm for providing guidance and focus of the learning and monitoring processes.

The following diagram depicts the Enhanced Rewind Paradigm for Learning.



III.B.3. The Concept of "Database of Intention"

Based on the proposed paradigm, a "database of intention" could be created based on the proposed paradigm for collecting and storing past actions and changes in environments, and learning from them. Such a database system focuses on the prediction of future intentions on users and their environmental changes based on their past behaviors. It augments human decision makers for taking planned actions. The system continuously monitors on going changes and continuously learning from the conceptually modeled dynamic real world environment for providing hints and alerts to human decision makers and planners. Specific constraints and conditions could be entered by human decision maker so that the system may learn from the past activities within the set of constraints provided by the human decision makers. The system is mainly for augmenting the human decision maker's intellectual capability for decision making under constraints. The "database of intention" expands the current database system into a new dimension for handling both static as well as dynamic data and information.

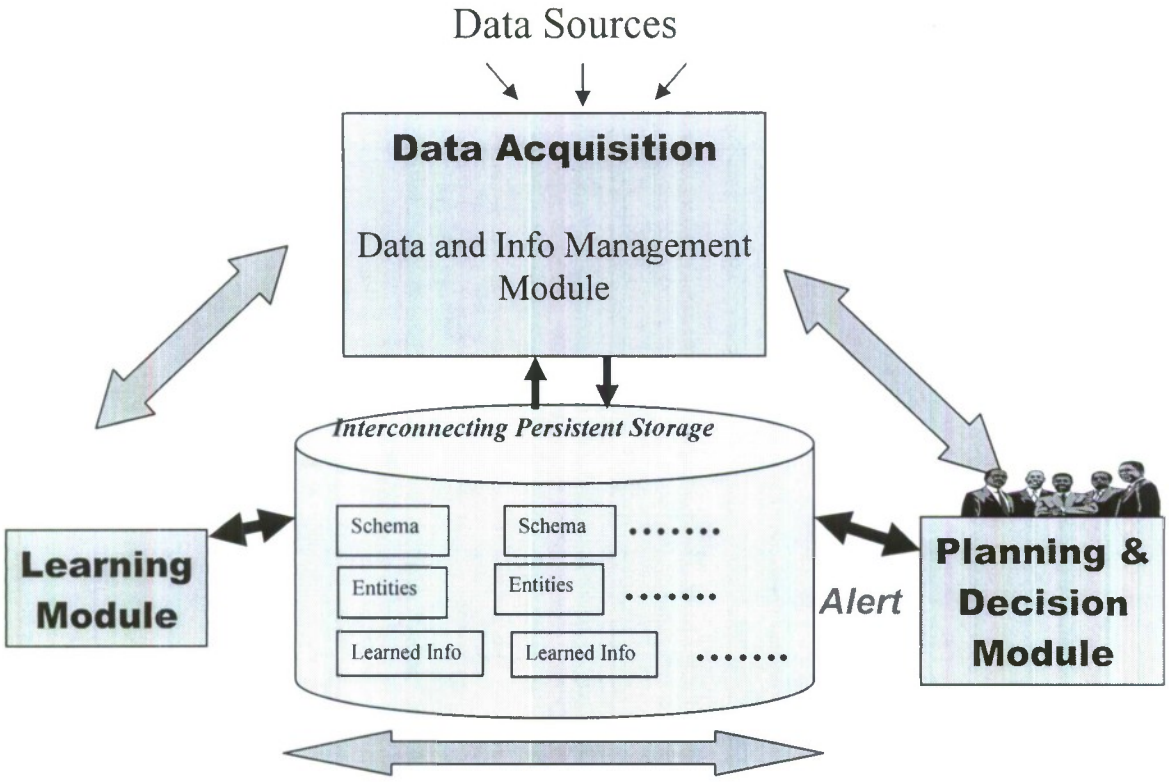
Based on the concept of "database of intension", a high level architecture for the next generation of database information technology is proposed. In this paper, the key technologies and components of the architecture are identified, illustrated, and discussed.
The proposed system architecture attempts to show the technical viability for the next generation of data and information system for providing a framework for research and development community for the development of prototypes, experimental test beds, and eventually the full operational system in the not too distant future.

III.B.4. Proposed System Architecture for the Next Generation of Information Systems

System architecture is proposed to provide a framework for the development of the next generation of information system. The proposed system architecture is based on the core technology of an interconnected mass persistent storage that can store and organize massive change data within the database and its domain data sources from the Internet and the real world. This mass persistent storage is managed by an executable conceptual model database management system so that it insures the consistency of data collection and data usage. The executable conceptual model database provides the flexibility for dynamically modeling the changing world by evolutionally changing the conceptual model as well as its associated data storage structures accordingly. The evolutional changes of the conceptual model are maintained in the persistent storage with their associated data to ensure that the changes in entity behavior as well as the traces and links that reveal "who, what, when, where and how" are directly associate with the correct version of the conceptual model in time.

The proposed architecture has three major interacting operational modules, namely, Data Acquisition Module, Learning Module, and Planning and Decision Making Module.
The following diagram depicts the proposed architecture.

System Architecture



III.B.4. Executable Conceptual Model Database Management System

The executable conceptual model data management system works more like an interpreter that directly interprets the actions specified in the conceptual model. It is different from the conventional database management system which uses the layered architecture for accomplishing data independence and the conceptual model is used as a database design tool for guiding the generation of different levels of schema. The executable conceptual model database management system is intended to execute the database management function directly through the conceptual model. The data collected in the interconnecting persistent storage can be linked directly with its corresponding elements in the conceptual model. This is necessary in order to make clear traces from applications through processes to entities and their instances. With the executable conceptual model database management system the instances of the entity are directly traceable to the entity and the entity is directly traceable to the conceptual model and applications. Therefore, the data in the persistent storage are closely coupled with the conceptual model. In a conventional database management system, the conceptual model is stored in a database design or CASE tools, the physical data in the database management system is isolated from the conceptual model and the application programs. Therefore, the architecture of conventional database management system created the difficulty for tracing storage data element back to its corresponding entity in the conceptual model. Reverse engineering has been used to accomplish this task without much success. The conventional architecture also introduced the difficult for evolutional changes in conceptual schema. When necessary, the database must be reorganized and regenerated which is a costly and time consuming task. Another important function of the active conceptual model is that it requires the handling of dynamic changes of the change of conceptual model itself in an evolutional manner. Schema evolution in a conventional database management system is a major difficulty that it requires the reorganization of the database with each modification of the conceptual model. By linking and integrating the major components together, the executable conceptual model data management system accomplishes traceability between application, conceptual schema, and storage structure.

IV. INVESTIGATION OF ADDITIONAL MATHEMATICAL MODELS/ALGORITHMS AND THEIR APPLICATIONS

Besides directly applying the mathematical framework we developed, we have also investigated other mathematical models and algorithms which are useful in a critical need of the Air Force – profiling problems of terrorists and malicious cyber transactions.

IV.1: The Usefulness of Mathematical Models and Algorithms for Profiling Problems

Suppose a group of objects is profiled by m attributes. In different applications, the objects can be different. For example, if we want to identify hijackers, then the objects would be all passengers; if we want to identify intruders to a network, then the objects would be all users of the network. In the following discussion, we will not distinguish these different scenarios since our method works in all situations.

To simplify our analysis, we assume that all the objects can be classified as *good*, the normal objects, and *bad*, the ones (hijackers or intruders) that we want to identify. In addition, we also assume that each attribute is a binary variable. That is, when applied to an object, each attribute returns either 0 or 1, indicating the attribute's suggestion on if the object is good or bad.

Since no attribute is perfect, some attributes could be wrong from time to time. The natural thing to do to increase accuracy is to use several attributes, instead of only one. However, when many attributes are developed, very often that some of them are redundant. Therefore, it is desirable to find the smallest set of attributes that can capture all the bad objects. At the same time, it is also important to reduce the disturbance to the good objects, which is more or less the same as reducing the total number of false alarms. In this project, we have formulated mathematical models to optimize both objectives.

We have formulated the problems as a family of the Weighted Set Covering Problem (WSCP). Both SCP and WSCP have been extensively studied in the past thirty years. It is known that SCP is NP-hard. Therefore, all problems mentioned above are also NP-hard. Unless P=NP, there is no polynomial time algorithm that can approximate SCP to any fixed factor. Consequently, such negative result also hold for all problems mentioned above.

On the positive side, the best known polynomial time algorithm that approximates WSCP is a greedy algorithm proposed by Chavatal. In this research project, we have developed several new algorithms including the extensions of Chavatal's algorithm. We then compare the performance of our algorithms with other existing algorithms and found out that our algorithms perform, in many cases, better than the existing algorithms!

In order to attack the profiling problem, we have studied extensively a particular mathematical function called, pseudo-Boolean function. Recently, we published several papers on new algorithms for pseudo-Boolean functions, which will be very useful in terrorist/malicious-cyber-transaction classification. We also started to look at social network analysis methods to see how to integrate them with the ER model.

VI.2. Pseudo-Boolean Functions

A special type of functions which are useful in the terrorist profiling problem is "pseudo-Boolean Function." The main question has been investigated by us is: how do we "learn" a pseudo-Boolean function efficiently? (A function is pseudo-Boolean if its variables take 0-1 values while the function may take any value. These functions could be used as a profiling tool.)

In [17], we present a new framework/method for learning pseudo-Boolean functions from training data. This framework is based on the observation that the training data can be seen as constraints on the possible candidate pseudo Boolean functions and that without any additional information, any of the pseudo Boolean function satisfying these constraints is equally likely. This framework was further studied in [15].

Another question we tried to answer was: if we have a (very complicated) pseudo-Boolean function, how can we get a simper yet good approximation? This is a very practical problem because the profiling functions are always very complex, meaning hard to compute. Therefore, having a good approximation would be useful. This problem is studied from different angles. In [19, 20], we studied the theoretical aspect, and we obtained closed form formulas (in many cases) for different kind of approximations. We also studied the algorithmic aspect of this problem and proposed efficient algorithms [14, 16, 17] for solving this problem.

V.3. Merging of Two Security Policies

Computer Security policies specify conditions for permissions to access various computer resources and information.

Merging two security policies is needed when two organizations together with their computer systems merge into one entity as in corporate business acquisition. In [21, 22] we propose a graph-theoretic method for merging the role/object hierarchies of two security policies. We formulate the merged security policy based on the {\it minor} relation in graph theory. Ideally, the merged role hierarchy should contain both the participating role hierarchies as graph minors, and similarly for the object hierarchy. We show that one can decide in polynomial time whether this ideal case is possible when the participating hierarchies are trees. In [23], we model a security system by a poset (partially ordered set) and we obtained optimal algorithms for merging various types of posets.

VI.4. New Formulation of the "Set Covering Problem" and New Algorithms

A profiling problem can be formulated as a set covering problem. To incorporate different requirements of a general profiling problem, we generalized the classical weighted set covering problem simultaneously in three directions. First, each numerical weight is replaced by a weighted set, which we call cost. Second, each element in the ground set is assigned a numerical weight. Third, the concept of a cover is relaxed to a partial cover that only needs to cover some percentage of the ground set, instead of the whole ground set. The last two generalizations have been studied in the literature, while the first is new. We propose a greedy algorithm to approximate this generalized problem and we establish an upper bound on the ratio of the greedy solution over the optimal solution. This bound is independent of the cost function, and it depends only on the total weight of the ground set. We prove that our bound is the best possible.

VI.5. Efficient Mathematical Algorithms for Learning, Classification, and data transmission security

We have investigated and obtained useful results when we studied the comparison of Greedy Strategies for Learning Markov Networks [24]. We have also studied and obtained interesting results in local soft belief updating for relational classification [25]. Both techniques are useful in learning and classifications (for example, for identification of potential terrorists).

Conclusions

We believe mathematical framework and models are very important to the Air Force and Department of Defense. If the Air Force can get a better understanding on fundamental principles of system architecture and the methodologies and techniques to apply system architecture concepts and tools correctly, the Air Force will be able to develop better systems and to maintain existing systems longer. Similarly, most organizations in the U.S. and the world should be able to design better products and systems (including its own organizational structure) based on the results reported in this project.

References

1. On The Complexity Of Rooted Trees And Hierarchies With Possible Applications To Organization Design And System Architectures (Peter Chen, G. Ding), WSEAS TRANSACTIONS on SYSTEMS, Issuc 3, Volume 5, March 2006, 625 -630.
2. Codynamics, "Introduction to the Basic Concepts of Complexity Science," in *http://www.codynamics.net/intro.htm*, 2004.
3. J. Collier, "Organized Complexity: Properties, Models, and Limits of Understanding," in *http://www.nu.ac.za/undphil/collier/papers/cuba-complexity.pdf*, 2004.
4. E. E. Olson, Glenda H. Eoyang, "Facilitating Organization Change: Lessons from Complexity Science," First ed: Jossey-Bass/Pfeiffer, February 7, 2001.
5. [M. Lissack, "Michael Lissack's Publications," in *http://lissack.com/writings/*, 2004.
6. T. Petzinger, "Complexity Reading List," in *http://www.petzinger.com/complexity.shtml*, 2004.
7. P. Chen, and Guoli Ding, "Unavoidable double-connected large graphs," *Discrete Mathematics*, 2004.
8. P. Chen, "A Preliminary Framework for Analyziing Critical Issues in Engineering Systems," prescnted at MIT Symposium on Engineering Systems, http://esd.mit.edu/symposium/pdfs/papers/chcn-abst.pdf, Cambridge, MA, 2004.
9. J. Moses, "Three Design Methodologies, Their Associated Organizational Structures and Their Relationships to Various Fields," presented at MIT Symposium on Engineering Systems, http://esd.mit.edu/symposium/pdfs/papers/moses.pdf, Cambride, MA, 2004.
10. P. Chen, "The entity-relationship model: Toward a unified view of data," *ACM Transactions on Database Systems*, vol. 1, 1976.
11. P. Chen, and Guoli Ding, "Generating r-regular graphs," *Discrete Applied Mathematics*, vol. 129, pp. 329-343, 2003.
12. P. Chen, and Guoli Ding, "The Best Expert Versus the Smartest Algorithm," *Theoretical Computer Science*, Vol. 332, No. 1-3, (2005), pp. 63-81.
13. T.C. Ting, Peter P. Chen, and Leah Wong (2007), "Architecture for Active Conceptual Modeling of Learning," Active Conceptual Modeling of Learning, Lecture Notes in Computer Science, Volume 4512, ACM-L 2006, edited by P.P. Chen and L.Y. Wong, Springer-Verlag Berlin Heidelberg, pp. 7 - 16.
14. Robert F. Lax, G. Ding, Peter Chen, and Jianhua Chen, "Asymptotic Behavior of Linear Approximations of Pseudo-Boolean Functions", *Proceedings of Taiwan Association for Artificial Intelligence International Conference, (Proc. the 10th Conference on Artificial Intelligence and Applications)*, December 2-3, 2005, Kaohsiung, Taiwan.
15. Jianhua Chen, G. Ding, Peter Chen, and Bob Lax, "Efficient Learning of Pseudo-Boolean Functions from Limited Training Data," *Lecture Notes in Computer Science, Volume 3488*, Page 323-331.
16. R. Lax, G. Ding, P. Chen and J. Chen, "Approximating Pseudo-Boolean Functions on Non-uniform domains," *Proc. International Joint Conference on Artificial Intelligence (IJCAI05)*, Page 1754-1755.
17. Jianhua Chen, Peter Chen, G. Ding, and Bob Lax, "A new method for learning pseudo-Boolean functions with applications in terrorists profiling ", *Proc. 2004 IEEE Conference on Cybernetics and Intelligent Systems, Volume 1*, (2005) 234- 239.
18. Robert F. Lax, G. Ding, Peter Chen, and Jianhua Chen, "Asymptotic Behavior of Linear Approximations of Pseudo-Boolean Functions," *Journal of Advanced Computational Intelligence and Intelligent Informatics, 11* (4) (2007) 403-409.
19. R.F. Lax, G. Ding, Jianhua Chen, and Peter P. Chen, "Formulas for Approximating Pscudo-Boolean Random Variables," *Discrete Applied Mathematics,* submitted.
20. R.F. Lax, G. Ding, Jianhua Chen, Peter Chen, and Brian D. Marx, "Transforms of pseudo-Boolcan random variables," *Discrete Applied Mathematics*, submitted.
21. G. Ding, J. Chen, P. Chen and R. Lax, "Graph-Theoretic Method for Merging Security System Specifications," *Information Sciences.*177 (10) (2007) 2152-2166.
22. Guoli Ding, Peter Chen and Steve Seiden, "Poset Merging with Applications to Databasc Security," *Discrete Mathematics & Theoretical Computer Science,* Submitted.
23. G. Ding, Peter Chen, "A greedy heuristic for a generalized set covering problem, *Discrete Mathematic,"* Submitted.

24. Robert Lax, Jianhua Chen, G. Ding, Peter P. Chen, and Brian Marx, "Comparison of Greedy Strategies for Learning Markov Networks of Treewidth k," *Proceedings of the 2007 International Conference on Machine Learning; Models, Technologies & Applications, MLMTA 2007*, June 25-28, 2007, Las Vegas Nevada, 294-301.
25. R.F. Lax, J. Chen, G. Ding, P. Chen, B.D. Marx, "Local soft belief updating for relational classification," *Proc. of the 17th International Symposium on Methodologies for Intelligent Systems (ISMIS 08),* York University, Toronto, Canada (May 20-23 2008).